On the Limits of the Principle of Sufficient Autonomy

CHRIS MILLS
University of Warwick

ABSTRACT

In his recent book, *Just Enough*, Liam Shields offers a novel defense of the Principle of Sufficient Autonomy. According to this principle, each citizen is owed 'enough' powers of deliberation and scope for decision-making as a matter of justice in order to satisfy our fundamental interest in acting and believing freely. In this article, I offer two objections against this view. The first objection challenges the plausibility of the principle. I argue that the principle that Shields derives from our interest in freedom will struggle to secure the proper protection for our capacities for autonomous behavior that many autonomy-minded liberals would expect the principle to provide. The second objection challenges the distinctiveness of the principle. I argue that Shields' defense cannot successfully dismiss all of its competitors and I offer an account of constitutive welfarism to illustrate this point.

1. INTRODUCTION

Theories of distributive justice cannot avoid questions concerning the value of choice and our capacities for autonomous decision-making in modern society. For example, we need to know how accessible our opportunities for valuable choice should be, how the opportunities to develop our decision-making capacities should be distributed between us, and how sensitive the distribution of goods and services should be to our individual choices. In his recent book, *Just Enough*, Liam Shields offers us a series of novel arguments that can help us answer these important questions.[1]

---

1   All in text references refer to this text.

Shields argues that we have a fundamental interest in enjoying the conditions of freedom; conditions under which our actions and beliefs can be considered freely taken (52). Each citizen requires sufficient autonomy in order to satisfy this interest because we each require some autonomy in order for our choices to be considered freely made by ourselves as agents. Our interest in freedom thus establishes the need for a *Principle of Sufficient Autonomy*.

This principle requires citizens to be: (a) well-informed, (b) able to give reasons for one's views, and (c) disposed to exchange reasons and participate in a public deliberative process with others (53). Fulfilling these three satiable conditions secures 'enough' autonomy, understood in terms of: (i) the citizen's powers of deliberation, and (ii) the scope of the decisions over which he or she decides (50).

This principle maintains a conception of autonomy that is framed in terms of an ideal of living one's life in accordance with one's own authentic judgements (47). It emphasizes the role that critical deliberation plays in establishing our capacity for self-rule (51) and concerns itself with avoiding threats to this capacity (especially threats of alien control that circumvent this capacity, such as coercion and manipulation) (48). This principle is relatively thin in its content and moderate in its demands.[2] It primarily focuses on establishing the conditions of authentic belief-formation and an ethos of well-informed and tolerant decision-making (53).

Shields' principle of sufficient autonomy is notable for two reasons. First, his defense of a sufficientarian principle of personal autonomy is distinctive. The relationship between personal autonomy and sufficiency has been defended in various ways. This relationship is most commonly cashed out in terms of option sets. For example, Joseph Raz (1986: 373-7) famously argues that personal autonomy requires agents to enjoy an adequate range of objectively valuable options. Gerald Dworkin (1988: 62-81) defends an adequacy limit on option sets according to a range of moral and rational considerations. Kerah Gordon-Solmon (2017) has recently offered a satisfaction-based defense of an adequacy limit on option sets grounded in the value of autonomy. However, this relationship has been explained in other ways. For example, Ben Colburn (2010: 89-92) defends a responsibility-sensitive threshold for autonomous capacities, below which we should not be held responsible for deficits in autonomy.

Although these arguments differ in their grounding or target, they all

---

  2   For example, Shields' principle has relatively little to say about the structure of autonomous motivation, limits on the validity of motivating factors, substantive constraints on option selection, relational constraints on the standing of autonomous agents, or limits on consent's role as a normative power.

serve to defend autonomy-sensitive principles of sufficiency as Shields understands the notion (i.e. as a shift or discontinuity in the rate of change in our marginal reasons to promote autonomy). Shields' argument is an important addition to this discourse. It distinguishes itself from preceding arguments by appealing to the Rawlsian higher-order interest in the social conditions of freedom as conditions capable of securing our freedom and equality as moral persons (52).

Second, the principle serves a pair of important functions in Shields' general argument for sufficientarianism. Shields primarily defends the principle as a central example of the indispensability of sufficientarianism to a sound and complete theory of distributive justice (26). In order to prove this, Shields identifies sufficientarian reasons as a distinctive type of non-instrumental, non-egalitarian, weighty, and satiable reason (44). He then sets out to prove that we cannot do without these distinctive reasons in the most plausible account of justice by showing that these reasons support principles that are more plausible than their rivals (17). Our reason to secure sufficient autonomy, and the principle that this reason supports, is a central example of this larger thesis.

The principle plays a further role as a significant bridging argument between Shields' general defense of sufficientarianism and his subsequent claims concerning the specific role of sufficientarian principles in education. He suggests that the fact that we owe sufficient autonomy to all should inform how we justify education for autonomy to groups who reject autonomy's value, how we educate in order to facilitate the discovery and development of talents, and how we conceive of the broader requirements of fair equality of opportunity (83).

The reasons why Shields' principle is notable are the very same reasons that motivate this study of his explanation and defense of the principle. His argument takes the form of a two-stage defense; first of the principle's content and then of its standing against competing principles. In this article, I offer a two-stage criticism that mirrors this strategy. First, I will object to the thinness of the principle and the protection of our capacity for autonomous behavior that it provides. I will then object to an important deficiency in his defense of the relative plausibility of his principle against its competitors.

2. THE FIRST OBJECTION – CAN THE PRINCIPLE ACHIEVE ITS AIMS?

Shields motivates his principle according to our interest in the conditions of freedom. This locates the principle of autonomy downstream from the interest in the conditions of freedom. This relationship explains why it is no objection to argue that the principle of sufficient autonomy does not provide us with enough autonomy to secure freedom. The proper role of the principle is not to secure all of the freedom that we might need. Rather, we can secure freedom through a number of principles, one of which must be the principle of sufficient autonomy.

However, we can legitimately object that the principle of sufficient autonomy, as it stands, does not provide us with all of the autonomy that our interest in the conditions of freedom should secure for us. This is the worry that I will press in this section. I argue that Shields' principle offers less protection for our capacities for autonomous behavior than it should, given that it is derived from our interest in freedom. This, in turn, leads us to question whether our interest in freedom is a satisfactory grounding for principles of autonomy.

This objection rests on the different ways in which our free and autonomous behavior can be hindered. Suppose that I sit down to write a short philosophy article. There are a variety of ways in which my decision to do so can fail to successfully translate into action through no fault of my own.

Shields' explanation of these failures explicitly focusses on the social conditions that influence our belief formation (53). He distinguishes between coercion (47) and other failures of self-direction that are the result of external forces acting on our beliefs (48). A long-suffering and frustrated neighbor who barges in to snatch up my notes and prevent me from writing is an example of the former. A manipulative neighbor who tricks me into giving up writing for the afternoon and going out for an enjoyable (though ultimately regrettable) walk instead is an example of the latter. Shields identifies both phenomena as possible threats to our autonomy because in both cases we are unfree to act or decide otherwise.

The principle of sufficient autonomy is intended to respond to these threats to our free and autonomous behavior by securing the social conditions under which our beliefs and actions can be considered freely taken (53). As we have seen, the principle focuses on the conditions of belief-formation (51), and in particular on an agent's powers of deliberation and the scope of the decisions over which he or she decides (50). This focus

generates the requirement that citizens are: (a) well-informed, (b) able to give reasons for their views, and (c) disposed to exchange reasons and participate in a public deliberative process with others. With this in mind, let us question whether Shields' distinction between coercion and other external threats acting on our beliefs is exhaustive and plausible. I suggest that it is not for the following reasons.

First, while it is true that my decision to write can be *circumvented* by some competing heteronomous motivation, it is not true that this motivation must stem from external forces. For example, I could deceive myself into over-estimating my writing ability and mistakenly lead myself into putting off my work for another day. Alternatively, a bout of hysteria or ambivalence may drive me to throw my notes out of the nearest window. In these instances, my initial intention to spend the afternoon writing has been foiled by a pernicious influence that leads to an inauthentic change of plan. I will subsequently become alienated from these decisions and come to authentically reject and regret them, just as I would if they were the product of external manipulation. Of course, it is well within my ability to change my mind as an autonomous agent and freely decide not to spend the afternoon writing. But there are troubling manifestations of this change of character that subvert my authentic will and are thus incompatible with my free and autonomous choice. Crucially, not all of these threats to my autonomy come from external sources, such as my neighbor.

Second, while it is true that my decision can successfully motivate me to act but that my motivation can still be subsequently *frustrated* by coercion (thus preventing me from acting), it is not true that coercion is the only phenomena that can frustrate my behavior in this way. For example, I may misplace my pen, fail to wrestle my notes out of the clutches of my pet, or be plunged into darkness thanks to a broken lightbulb. In these instances, no other agent has frustrated my autonomous decision to write. Rather, frustration is the result of simply lacking the option to perform the act that I had autonomously chosen to perform. It is frustration, rather than coercion, that prevents me from behaving authentically in these cases. Frustration can occur by either natural accident or inter-personal sabotage. While extreme forms of frustration should not concern us (e.g. the irrational desire to perform the impossible), some forms of non-coercive frustration clearly threaten our free and autonomous decision-making.

Circumvention and frustration come apart in a similar manner to Shields' own distinction between coercion and other external threats to our autonomy. Indeed, frustration will similarly occur in the absence of

circumvention as the latter precedes the former in the chain of action. However, both circumvention and frustration are broader than Shields' categories of coercion and other external threats. As a result, if you agree that my distinction provides a more plausible and comprehensive explanation of the various threats to our autonomy that we face, then you may worry that Shields' principle of sufficient autonomy does a poor job of protecting our authentic decisions against internal threats (such as self-deception) or non-interpersonal external threats (such as bad luck). As a result, the principle appears to do a poor job at protecting our autonomy from threats that Shields ignores.

This objection is similar in form to Shields' own objection against John Rawls. In his discussion of Rawls' argument from the interest in freedom, Shields notes a possible ambiguity. Shields interprets Rawls' argument to support the possibility of achieving a sufficient level of autonomy as one important option that should be open to citizens. This is too small a commitment from Rawls. Our interest in freedom does not merely require the possibility of achieving sufficient autonomy, but rather the *actual* achievement of sufficient autonomy (55). Without the actual achievement of sufficient autonomy, we cannot know that each citizen's decision whether or not to live an autonomous life is itself free. Given that our interest in freedom suggests that we should strive to make sure that our adoption of an autonomous lifestyle is itself freely chosen, we require a larger commitment from Rawlsians in their defense of sufficient autonomy.

Shields' defense of the principle of sufficient autonomy is guilty of the same failing for which he dismisses Rawls' argument; at best, Shields' argument is necessary but not sufficient for establishing the conclusion that he wishes to draw. While it is true that the threats that he identifies are likely threats to autonomy, there are other threats to autonomy that should plausibly be recognized as contrary to our interest in freedom.

Shields may respond to this objection in one of two ways: he may concede by fleshing out his argument to encompass further types of threat. Alternatively, he may resist by rejecting the notion that non-interpersonal threats (such as internal threats or accidents) threaten our freedom. According to this response, he has not mistakenly ignored a range of likely threats. Rather, freedom is a question of interpersonal interactions not opportunities for autonomous action. For this reason, non-interpersonal threats should not be covered by a principle of sufficient autonomy that is grounded in our interests in freedom. Those of us who are concerned with protecting further opportunities for autonomous action can look to other compatible reasons to promote autonomy (45), but Shields' focus on securing enough autonomy results from his core sufficientarian reason,

and this reason only concerns interpersonal threats (that is, threats to our interest in freedom from other agents). Therefore, Shields' principle is rightly insensitive to non-interpersonal concerns.

This response is important because it shifts our gaze to the deeper question concerning the justification of his principle. Sceptics may meet him here by denying the downstream relationship and arguing that our autonomy is threatened by more than a mere loss of freedom and therefore Shields' principle is incorrectly justified. Both circumvention and frustration undermine our autonomy and, by doing so, prevent us from enjoying the conditions of freedom in line with our authentic conception of the good. We cannot pursue the opportunities afforded to us in line with our authentic wishes if we are constantly self-sabotaging or suffering from a pronounced mismatch between our preferences and our option set.[3] Therefore, his principle is too thin because its justification is wrong. Autonomy tells us which freedoms matter, not the other way around.

Sceptics may conclude that Shields' principle only offers us an incomplete defense of our autonomy because he derives it from an interest in interpersonal considerations of freedom. This conclusion explains why Shields' principle is likely to be attractive to Rawlsians (who may share the same conception and weighting of our interest in freedom) but unattractive to other autonomy-minded liberals who worry about a broader set of threats to our autonomy.[4] Seen in this light, the feature that makes Shields' argument distinctive is also a limitation.

3. THE SECOND OBJECTION – IS THE PRINCIPLE PREFERABLE TO ITS COMPETITORS?

Having questioned the content of Shields' principle, I now turn to his defense of its relative plausibility against competing principles. Shields' rebuttal of his competitors first rejects rival distributive schemes of

---

3   These issues combine in cases of adaptive preferences. As fans of famous fables involving foxes will know, the fox that cannot reach the nearby bunch of grapes may mistakenly conclude that they must be sour, and thus undesirable. Of course, adaption cases need not be so far-fetched. But even mundane versions of this phenomena can threaten our autonomy. Crucially, the grapes need not be lifted out of reach by a devious neighbour for our out-of-character decision to override our preferences. Rather, the mere absence of an option can trouble autonomous decision-making. For discussion, see Elster (1983); Colburn (2011); Christman (2014); Stoljar (2014); Cudd (2014).

4   As a reviewer helpfully suggests, the Rawlsian's support will hinge on how they view Shields' interpretation of the higher-order interest in the social conditions of freedom. A less-relational reading of this interest will make Rawlsians more likely share my concerns over Shields' argument. If this is the case, then Shields' view is even less attractive.

autonomy (equality, maximization, and priority) and then rejects rival accounts of the relationship between autonomy and other distributive values (instrumentalism). After quickly dismissing egalitarian and maximizing principles, Shields focuses much of his argument against two main competitors:

1) *Uniform Prioritarianism* - that those who have the least autonomy should be prioritized with no 'shift' or discontinuity in the rate of change in our marginal reasons to promote autonomy.

Shields rejects uniform prioritarianism because: (a) it cannot plausibly explain qualitatively different autonomy disadvantages, and (b) it requires a non-arbitrary measure of autonomy that allows us to make fine-grained distinctions at all levels of the distribution scale. If prioritarianism is to distinguish itself from sufficientarianism then it must provide a uniform metric for the distribution of autonomy that avoids appeal to a threshold. Without this threshold, our metric must provide a plausible explanation of how we are better or worse off in terms of autonomy at points all along the distribution scale. Shields is rightly skeptical that such a measure exists (69).[5]

2) *Instrumental Welfarism* - that securing sufficient autonomy is an important demand of justice iff it has great effects on the more fundamental value of welfare.

Shields rejects instrumental welfarism with a pair of counter-examples (74). These examples aim to show that fully instrumental accounts of the value of autonomy fail to capture all of our intuitions about the non-instrumental role that the value of autonomy plays both in our lives and in a complete and sound theory of distributive justice. These examples suggest that welfarists who believe that autonomy holds purely instrumental value are committed to implausible conclusions, such as the permissibility of bypassing our deliberative capacities or shaping our ambitions to ensure that citizens live good lives. Even if citizens are guaranteed a well lived life, we should suspect that something important is missing in such cases.

However, crucially, Shields fails to dismiss accounts of how welfare is partly constituted by autonomy and so is derivatively but non-instrumentally significant (71). We might call such views *Constitutive Welfarism*[6] It is true that constitutive welfarism is compatible with Shields'

---

5   For more detail on this difficulty, see Blake (2001: 269).

6   I do not intend constitutive welfarism as the only non-instrumental justification of the principle of sufficient autonomy (consider, for example, a Kantian justification of the principle). However, I do intend it as one in a small possible set of justifications that can solve Shields' objections to uniform prioritarianism in a way that doesn't fall foul of his objections to instrumental welfarism.

claim that autonomy matters non-instrumentally. These views do not deny that there is a non-instrumental principle of autonomy and thus do not challenge Shields' arguments concerning the indispensability of the principle of sufficient autonomy. However, Shields must still dismiss these views because, although they are in agreement with his stance against pure instrumentalism, they compromise his arguments against uniform prioritarianism. They do this by explaining autonomy's non-instrumental value in terms of the constitutive role that autonomssy plays in promoting good lives. This autonomy-sensitive notion of welfare, in turn, can provide a compelling metric to measure our access to autonomy that is otherwise missing from uniform prioritarianism.

To see this, assume that you agree that the value of autonomy plays a constitutive role in living a good life. If you believe this to be true, then you believe that some plausible account of well-being can explain the value of autonomy in a non-instrumental manner. For example, you might agree with perfectionists Joseph Raz (1986: 391) and Steven Wall (1998: 164-182) that the social forms of a liberal society require those who seek to live a good life in such a society to possess enough personal autonomy to make sense of the valuable options available to them. Alternatively, you might agree with anti-perfectionists Will Kymlicka (1989: 10-13) and Ronald Dworkin (2000: 267-274) that authentic endorsement has a necessary role to play in explaining the value of a life well lived. These arguments (and others) open up the conceptual space for a principle that ties notions of autonomy and authenticity (similar to those favored by Shields) to well-being in a non-instrumental fashion. These principles will capture all of our intuitions about the non-instrumental role that the value of autonomy plays both in our lives and in a complete and sound theory of distributive justice. This explains why constitutive welfarism cannot be dismissed by Shields' counter-examples.

Troublingly, these arguments allow uniform prioritarians to appeal to the value of well-being (suitably conceived) as a non-arbitrary metric of autonomy in order to defend their view against Shields' objections to the position. If you are a uniform prioritarian about welfare, then constitutive welfarism allows you to: (a) distinguish between qualitatively different disadvantages, and (b) appeal to some reasonably fine-grained metric of well-being in order to distribute autonomy without necessarily appealing to sufficientarian reasons.

Not only does constitutive welfarism evade the theoretical traps that Shields lays out for his competitors, it may provide some with a more plausible explanation of the role that autonomy should play in a just society

than Shields' own narrower Rawlsian framework. As we saw in §2, the protection for autonomous behavior provided by Shields' principle of sufficient autonomy is limited by its grounding in our interest in freedom. I suggested that some autonomy-minded liberals might be disappointed by this. This disappointment could be undercut if there are no other plausible alternative justifications for the principle. If this were true, then Shields may offer his principle as the only show in town. However, constitutive welfarism offers us an alternative show. This justification does not tie our capacities for autonomy to our interest in freedom, but instead ties both our capacities for autonomy and freedom to our interest in living good lives. While such a view may trouble Rawlsians, Shields must provide further counter-arguments to reject this competitor.[7]

4. CONCLUSION

I have offered two brief objections to Shields' novel defense of the principle of sufficient autonomy. The first objection calls for Shields to broaden his principle in order to more robustly satisfy the interest in freedom that motivates his arguments, and to ultimately reconsider this motivation. The second objection calls for Shields to provide further justification for his principle in a manner that is more nuanced and more sensitive to competing positions. Only an argument that satisfies these challenges will prove the indispensability of his principle of sufficient autonomy to a sound and complete theory of distributive justice.

BIBLIOGRAPHY

Blake, M., 2001: "Distributive Justice, State Coercion, and Autonomy", *Philosophy and Public Affairs* 30: 257-296.

Christman, J., 2014: "Coping or Oppression: Autonomy and Adaption to Circumstance", in *Autonomy, Oppression, and Gender*, ed. A. Veltman and M. Piper, 201-226, Oxford: Oxford University Press.

Colburn, B., 2010: *Autonomy and Liberalism*, New York: Routledge.

— 2011: "Autonomy and Adaptive Preferences", *Utilitas* 23: 52-71.

---

7   For example, the door remains open for Shields to argue that constitutive welfarism must still secure a shift, either because the most plausible account of well-being requires us to have 'enough' autonomy, or at one level removed, because that account allows us to live a 'good enough' life. Neither of these threshold arguments are out of the question for constitutive welfarism. However, this strategy requires Shields to accept another (non-Rawlsian) defence of a competing principle of sufficient autonomy. Accepting the existence of a valid competitor will lessen the importance of his distinctive argument to the broader case for sufficientarianism

Cudd, A.E., 2014: "Adaptions to Oppression: Preference, Autonomy, and Resistance", in *Personal Autonomy and Social Oppression: Philosophical Perspectives*, ed. M. Oshana, 142-160, New York: Routledge.

Dworkin, G., 1988: *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press.

Dworkin, R., 2000: *Sovereign Virtue: The Theory and Practice of Equality*, Cambridge Mass.: Harvard University Press.

Elster, J., 1983: *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge: Cambridge University Press.

Gordon-Solmon, K., 2017: "Why More Choice is Sometimes Worse than Less", *Law and Philosophy* 36: 25-44.

Kymlicka, W., 1989: *Liberalism, Community, and Culture*, Oxford: Clarendon Press.

Raz, J., 1986: *The Morality of Freedom*, Oxford: Clarendon Press.

Shields, L., 2016: *Just Enough: Sufficiency as a Demand of Justice*, Edinburgh: Edinburgh University Press.

Stoljar, N., 2014: "Autonomy and Adaptive Preference Formation", in *Autonomy, Oppression, and Gender*, ed. A. Veltman and M. Piper, 227-253, Oxford: Oxford University Press.

Wall, S., 1998: *Liberalism, Perfectionism, and Restraint*, Cambridge: Cambridge University Press.